

SEQUENCE LISTING

<110> Whitney, Mike
Xanthopoulos, Kleanthis
Nelson, David
Negulescu, Paul
Craig, Frank
Foulkes, J. Gordon

<120> METHODS AND COMPOSITIONS FOR SENSITIVE
AND RAPID, FUNCTIONAL IDENTIFICATION OF GENOMIC
POLYNUCLEOTIDES AND USE FOR CELLULAR ASSAYS IN DRUG
DISCOVERY

<130> 08366/026001

<140> 09/047,862

<141> 1998-03-25

<150> 09/021,974

<151> 1998-02-11

<150> 08/719,697

<151> 1996-09-26

<160> 15

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 795

<212> DNA

<213> Escherichia coli

<400> 1

atgagtcacc	cagaaacgct	ggtgaaagta	aaagatgctg	aagatcagtt	gggtgcacga	60
gtgggttaca	togaactgga	tctcaacagc	ggtaagatcc	ttgagagttt	tcgccccgaa	120
gaacgttttc	caatgatgag	cactttttaa	gttctgctat	gtggcgcggt	attatcccgt	180
gttgacgccg	ggcaagagca	actcggtcgc	cgcatacact	attctcagaa	tgacttggtt	240
gagtactcac	cagtcacaga	aaagcatctt	acggatggca	tgacagtaag	agaattatgc	300
agtgtcgcca	taaccatgag	tgataacact	gcggccaact	tacttctgac	aacgatcgga	360
ggaccgaagg	agctaaccgc	ttttttgcac	aacatggggg	atcatgtaac	tcgccttgat	420
cgttgggaac	cggagctgaa	tgaagccata	ccaaacgacg	agcgtgacac	cacgatgcct	480
gcagcaatgg	caacaacggt	gcgcaacta	ttaactggcg	aactacttac	tctagcttcc	540
cggcaacaat	taatagactg	gatggaggcg	gataaagttg	caggaccact	tctgcgctcg	600
gcccttccgg	ctggctggtt	tattgctgat	aaatctggag	ccggtgagcg	tgggtctcgc	660
ggtatcattg	cagcactggg	gccagatggt	aagccctccc	gtatcgtagt	tatctacacg	720
acggggagtc	aggcaactat	ggatgaacga	aatagacaga	tcgctgagat	aggtgcctca	780
ctgattaagc	attgg					795

<210> 2

<211> 858

<212> DNA

<213> Escherichia coli

<400> 2

atgagaattc	aacatttccg	tgctgccctt	attccctttt	ttgcggcatt	ttgccttcc	60
gtttttggtc	accagaaac	gctggtgaaa	gtaaaagatg	ctgaagatca	gttgggtgca	120
cgagtgggtt	acatcgaact	ggatctcaac	agcggtaaga	tccttgagag	ttttcgcccc	180

097214-012601

gaagaacggt	ttccaatgat	gaatctttt	aaagttctgc	tatgtggcgc	ggatctatcc	240
cgtgttgacg	ccgggcaaga	gcaactcggt	cgccgcatac	actattctca	gaatgacttg	300
ggtgagtact	caccagtcac	agaaaagcat	cttacggatg	gcatgacagt	aagagaatta	360
tgcagtgtctg	ccataaccat	gagtataaac	actgcgccca	acttacttct	gacaacgatc	420
ggaggaccga	aggagctaac	cgcttttttg	cacaacatgg	gggatcatgt	aactcgcctt	480
gatcggtggg	aaccggagct	gaatgaagcc	ataccaaacg	acgagcgtga	caccacgatg	540
cctgcagcaa	tggcaacaac	gttgcgcaaa	ctattaactg	gcgaactact	tactctagct	600
tcccggcaac	aattaataga	ctggatggag	goggataaag	ttgcaggacc	acttctgcgc	660
tcggcccttc	cggtcggtg	gtttattgct	gataaatctg	gagccggtga	gcgtgggtct	720
cgcggtatca	ttgcagcact	ggggccagat	ggtaagccct	cccgatcgt	agttatctac	780
acgacgggga	gtcaggcaac	tatggatgaa	cgaaatagac	agatcgctga	gataggtgcc	840
tcactgatta	agcattgg					858

<210> 3

<211> 843

<212> DNA

<213> Escherichia coli

<400> 3

aagctttttg	cagaagctca	gaataaaocg	aactttccgg	gtaccaccat	ggggcaccca	60
gaaacgctgg	tgaaagtaaa	agatgctgaa	gatcagttgg	gtgcacgagt	gggttacatc	120
gaactggatc	tcaacagcgg	taagatcctt	gagagttttc	gccccgaaga	acgttttcca	180
atgatgagca	cttttaaagt	tctgctatgt	ggcgcggtat	tatcccgatga	tgacgcgggg	240
caagagcaac	tcggtcgccg	catacactat	tctcagaatg	acttggttga	gtactcacca	300
gtcacagaaa	agcatcttac	ggatggcatg	acagtaagag	aattatgcag	tgctgccata	360
accatgagtg	ataacactgc	ggccaactga	cttctgacaa	cgatcggagg	accgaaggag	420
ctaaccgctt	ttttgcacaa	catgggggat	catgtaaact	gccttgatca	ttgggaaccg	480
gagctgaatg	aagccatacc	aaacgcagag	cgtgacacca	cgatgcctgt	agcaatggca	540
acaacggttg	gcaaaactatt	aactggcgaa	ctacttactc	tagcttcccg	gcaacaatta	600
atagactgga	tggaggcgga	taaagttgca	ggaccacttc	tgcgctcggc	ccttcgggct	660
ggctgggttta	ttgctgataa	atctggagcc	ggtgagcgtg	ggtctcgcg	tatcattgca	720
gcaactggtg	cagatggtaa	gccctcccg	atcgtagtta	tctacacgac	ggggagtcag	780
gcaactatgg	atgaacgaaa	tagacagatc	gctgagatag	gtgcctcact	gattaagcat	840
tgg						843

<210> 4

<211> 792

<212> DNA

<213> Escherichia coli

<400> 4

atggacccag	aaacgctggt	gaaagtaaaa	gatgctgaag	atcagttggg	tgcacgagtg	60
ggttacatcg	aactggatct	caacagcggg	aagatccttg	agagttttcg	ccccgaagaa	120
cgttttccaa	tgatgagcac	ttttaaaagt	ctgctatgtg	gcgcgggtatt	atcccgtatt	180
gacgccgggc	aagagcaact	cggtcgccgc	atacactatt	ctcagaatga	cttggttgag	240
tactcaccag	tcacagaaaa	gcatcttacg	gatggcatga	cagtaagaga	attatgcagt	300
gctgccataa	ccatgagtga	taacactgcg	gccaacttac	ttctgacaa	gatcggagga	360
ccgaaggagc	taaccgcttt	tttgacaa	atgggggatc	atgtaactcg	ccttgatcat	420
tgggaaccgg	agctgaatga	agccatacca	aacgacgagc	gtgacaccac	gatgcctgta	480
gcaatggcaa	caacgttgcy	caaactatta	actggcgaa	tacttactct	agcttcccg	540
caacaattaa	tagactggat	ggaggcggat	aaagttgcag	gaccacttct	gcgctcggcc	600
cttcgggctg	gctgggttat	tgctgataaa	tctggagccg	gtgagcgtgg	gtctcgggt	660
atcattgcag	cactggggcc	agatggtaag	ccctcccgta	tcgtagttat	ctacacgagc	720
gggagtcagg	caactatgga	tgaacgaaat	agacagatcg	ctgagatagg	tgctcactg	780
attaagcatt	gg					792

<210> 5

<211> 786

<212> DNA

<213> Bacillus licheniformis

<400> 5
 atgaaagatg attttgcaaa acttgaggaa caatttgatg caaaactcgg gatctttgca 60
 ttggatacag gtacaaaccg gacggtagcg tatcggccgg atgagcggtt tgcttttgct 120
 togacgatta aggctttaac tgtaggcgtg cttttgcaac agaaatcaat agaagatctg 180
 aaccagagaa taacatatac acgtgatgat cttgtaaact acaacccgat tacggaaaag 240
 cacgttgata cgggaatgac gctcaaagag cttgcggatg cttcgcttcg atatatgtgac 300
 aatgcggcac agaattctcat tcttaaacia attggcggac ctgaaagttt gaaaaaggaa 360
 ctgaggaaga ttggtgatga ggttacaaat cccgaacgat tcgaaccaga gttaaatagaa 420
 gtgaatccgg gtgaaactca ggataccagt acagcaagag cacttggtcac aagccttcga 480
 gcctttgctc ttgaagataa acttccaagt gaaaaacggc agcttttaaat cgattggatg 540
 aaacgaaata ccaactggaga cgccttaatc cgtgccggag cggcatcata tggaacccgg 600
 aatgacattg ccatcatttg gccgccaaaa ggagatcctg tcggtgtgcc ggacggttgg 660
 gaagtggctg ataaaactgt tcttgcaagta ttatccagca gggataaaaa ggacgccaag 720
 tatgatgata aacttattgc agaggcaaca aaggtggtaa tgaaagcctt aaacatgaac 780
 ggcaaa 786

<210> 6
 <211> 265
 <212> PRT
 <213> Escherichia coli

<400> 6
 Met Ser His Pro Glu Thr Leu Val Lys Val Lys Asp Ala Glu Asp Gln
 1 5 10 15
 Leu Gly Ala Arg Val Gly Tyr Ile Glu Leu Asp Leu Asn Ser Gly Lys
 20 25 30
 Ile Leu Glu Ser Phe Arg Pro Glu Glu Arg Phe Pro Met Met Ser Thr
 35 40 45
 Phe Lys Val Leu Leu Cys Gly Ala Val Leu Ser Arg Val Asp Ala Gly
 50 55 60
 Gln Glu Gln Leu Gly Arg Arg Ile His Tyr Ser Gln Asn Asp Leu Val
 65 70 75 80
 Glu Tyr Ser Pro Val Thr Glu Lys His Leu Thr Asp Gly Met Thr Val
 85 90 95
 Arg Glu Leu Cys Ser Ala Ala Ile Thr Met Ser Asp Asn Thr Ala Ala
 100 105 110
 Asn Leu Leu Leu Thr Thr Ile Gly Gly Pro Lys Glu Leu Thr Ala Phe
 115 120 125
 Leu His Asn Met Gly Asp His Val Thr Arg Leu Asp Arg Trp Glu Pro
 130 135 140
 Glu Leu Asn Glu Ala Ile Pro Asn Asp Glu Arg Asp Thr Thr Met Pro
 145 150 155 160
 Ala Ala Met Ala Thr Thr Leu Arg Lys Leu Leu Thr Gly Glu Leu Leu
 165 170 175
 Thr Leu Ala Ser Arg Gln Gln Leu Ile Asp Trp Met Glu Ala Asp Lys
 180 185 190
 Val Ala Gly Pro Leu Leu Arg Ser Ala Leu Pro Ala Gly Trp Phe Ile
 195 200 205
 Ala Asp Lys Ser Gly Ala Gly Glu Arg Gly Ser Arg Gly Ile Ile Ala
 210 215 220
 Ala Leu Gly Pro Asp Gly Lys Pro Ser Arg Ile Val Val Ile Tyr Thr
 225 230 235 240
 Thr Gly Ser Gln Ala Thr Met Asp Glu Arg Asn Arg Gln Ile Ala Glu
 245 250 255
 Ile Gly Ala Ser Leu Ile Lys His Trp
 260 265

<210> 7
 <211> 285

09772144 012601

<212> PRT
<213> Escherichia coli

<400> 7

Arg	Ile	Gln	His	Phe	Arg	Val	Ala	Leu	Ile	Pro	Phe	Phe	Ala	Ala	Phe
1				5					10					15	
Cys	Leu	Pro	Val	Phe	Gly	His	Pro	Glu	Thr	Leu	Val	Lys	Val	Lys	Asp
			20					25					30		
Ala	Glu	Asp	Gln	Leu	Gly	Ala	Arg	Val	Gly	Tyr	Ile	Glu	Leu	Asp	Leu
		35					40					45			
Asn	Ser	Gly	Lys	Ile	Leu	Glu	Ser	Phe	Arg	Pro	Glu	Glu	Arg	Phe	Pro
	50					55					60				
Met	Met	Ser	Thr	Phe	Lys	Val	Leu	Leu	Cys	Gly	Ala	Val	Leu	Ser	Arg
65					70					75					80
Val	Asp	Ala	Gly	Gln	Glu	Gln	Leu	Gly	Arg	Arg	Ile	His	Tyr	Ser	Gln
			85					90						95	
Asn	Asp	Leu	Val	Glu	Tyr	Ser	Pro	Val	Thr	Glu	Lys	His	Leu	Thr	Asp
		100						105					110		
Gly	Met	Thr	Val	Arg	Glu	Leu	Cys	Ser	Ala	Ala	Ile	Thr	Met	Ser	Asp
		115					120					125			
Asn	Thr	Ala	Ala	Asn	Leu	Leu	Leu	Thr	Thr	Ile	Gly	Gly	Pro	Lys	Glu
	130					135					140				
Leu	Thr	Ala	Phe	Leu	His	Asn	Met	Gly	Asp	His	Val	Thr	Arg	Leu	Asp
145					150					155					160
Arg	Trp	Glu	Pro	Glu	Leu	Asn	Glu	Ala	Ile	Pro	Asn	Asp	Glu	Arg	Asp
			165						170						175
Thr	Thr	Met	Pro	Ala	Ala	Met	Ala	Thr	Thr	Leu	Arg	Lys	Leu	Leu	Thr
		180						185					190		
Gly	Glu	Leu	Leu	Thr	Leu	Ala	Ser	Arg	Gln	Gln	Leu	Ile	Asp	Trp	Met
		195					200					205			
Glu	Ala	Asp	Lys	Val	Ala	Gly	Pro	Leu	Leu	Arg	Ser	Ala	Leu	Pro	Ala
	210					215					220				
Gly	Trp	Phe	Ile	Ala	Asp	Lys	Ser	Gly	Ala	Gly	Glu	Arg	Gly	Ser	Arg
225					230					235					240
Gly	Ile	Ile	Ala	Ala	Leu	Gly	Pro	Asp	Gly	Lys	Pro	Ser	Arg	Ile	Val
			245						250					255	
Val	Ile	Tyr	Thr	Thr	Gly	Ser	Gln	Ala	Thr	Met	Asp	Glu	Arg	Asn	Arg
		260						265					270		
Gln	Ile	Ala	Glu	Ile	Gly	Ala	Ser	Leu	Ile	Lys	His	Trp			
	275						280					285			

<210> 8

<211> 265

<212> PRT

<213> Escherichia coli

<400> 8

Met	Gly	His	Pro	Glu	Thr	Leu	Val	Lys	Val	Lys	Asp	Ala	Glu	Asp	Gln
1				5					10					15	
Leu	Gly	Ala	Arg	Val	Gly	Tyr	Ile	Glu	Leu	Asp	Leu	Asn	Ser	Gly	Lys
		20						25					30		
Ile	Leu	Glu	Ser	Phe	Arg	Pro	Glu	Glu	Arg	Phe	Pro	Met	Met	Ser	Thr
		35					40					45			
Phe	Lys	Val	Leu	Leu	Cys	Gly	Ala	Val	Leu	Ser	Arg	Asp	Asp	Ala	Gly
	50					55					60				
Gln	Glu	Gln	Leu	Gly	Arg	Arg	Ile	His	Tyr	Ser	Gln	Asn	Asp	Leu	Val
65					70					75					80
Glu	Tyr	Ser	Pro	Val	Thr	Glu	Lys	His	Leu	Thr	Asp	Gly	Met	Thr	Val
			85					90					95		
Arg	Glu	Leu	Cys	Ser	Ala	Ala	Ile	Thr	Met	Ser	Asp	Asn	Thr	Ala	Ala

Asn	Leu	Leu	100	Thr	Thr	Ile	105	Gly	Gly	Pro	Lys	Glu	Leu	110	Thr	Ala	Phe
			115				120							125			
Leu	His	Asn	Met	Gly	Asp	His	Val	Thr	Arg	Leu	Asp	His	Trp	Glu	Pro		
			130				135							140			
Glu	Leu	Asn	Glu	Ala	Ile	Pro	Asn	Asp	Glu	Arg	Asp	Thr	Thr	Met	Pro		
						150					155				160		
Val	Ala	Met	Ala	Thr	Thr	Leu	Arg	Lys	Leu	Leu	Thr	Gly	Glu	Leu	Leu		
						165					170				175		
Thr	Leu	Ala	Ser	Arg	Gln	Gln	Leu	Ile	Asp	Trp	Met	Glu	Ala	Asp	Lys		
						180					185				190		
Val	Ala	Gly	Pro	Leu	Leu	Arg	Ser	Ala	Leu	Pro	Ala	Gly	Trp	Phe	Ile		
						195					200				205		
Ala	Asp	Lys	Ser	Gly	Ala	Gly	Glu	Arg	Gly	Ser	Arg	Gly	Ile	Ile	Ala		
						210					215				220		
Ala	Leu	Gly	Pro	Asp	Gly	Lys	Pro	Ser	Arg	Ile	Val	Val	Ile	Tyr	Thr		
						225					230				235		
Thr	Gly	Ser	Gln	Ala	Thr	Met	Asp	Glu	Arg	Asn	Arg	Gln	Ile	Ala	Glu		
						245					250				255		
Ile	Gly	Ala	Ser	Leu	Ile	Lys	His	Trp									
						260					265						

<210> 9

<211> 264

<212> PRT

<213> Escherichia coli

<400> 9

Met	Asp	Pro	Glu	Thr	Leu	Val	Lys	Val	Lys	Asp	Ala	Glu	Asp	Gln	Leu		
						5			10					15			
Gly	Ala	Arg	Val	Gly	Tyr	Ile	Glu	Leu	Asp	Leu	Asn	Ser	Gly	Lys	Ile		
						20			25					30			
Leu	Glu	Ser	Phe	Arg	Pro	Glu	Glu	Arg	Phe	Pro	Met	Met	Ser	Thr	Phe		
						35			40					45			
Lys	Val	Leu	Leu	Cys	Gly	Ala	Val	Leu	Ser	Arg	Ile	Asp	Ala	Gly	Gln		
						50			55			60					
Glu	Gln	Leu	Gly	Arg	Arg	Ile	His	Tyr	Ser	Gln	Asn	Asp	Leu	Val	Glu		
						65			70			75			80		
Tyr	Ser	Pro	Val	Thr	Glu	Lys	His	Leu	Thr	Asp	Gly	Met	Thr	Val	Arg		
						85			90					95			
Glu	Leu	Cys	Ser	Ala	Ala	Ile	Thr	Met	Ser	Asp	Asn	Thr	Ala	Ala	Asn		
						100			105					110			
Leu	Leu	Leu	Thr	Thr	Ile	Gly	Gly	Pro	Lys	Glu	Leu	Thr	Ala	Phe	Leu		
						115			120					125			
His	Asn	Met	Gly	Asp	His	Val	Thr	Arg	Leu	Asp	His	Trp	Glu	Pro	Glu		
						130						140					
Leu	Asn	Glu	Ala	Ile	Pro	Asn	Asp	Glu	Arg	Asp	Thr	Thr	Met	Pro	Val		
						145					155				160		
Ala	Met	Ala	Thr	Thr	Leu	Arg	Lys	Leu	Leu	Thr	Gly	Glu	Leu	Leu	Thr		
						165					170				175		
Leu	Ala	Ser	Arg	Gln	Gln	Leu	Ile	Asp	Trp	Met	Glu	Ala	Asp	Lys	Val		
						180			185					190			
Ala	Gly	Pro	Leu	Leu	Arg	Ser	Ala	Leu	Pro	Ala	Gly	Trp	Phe	Ile	Ala		
						195			200					205			
Asp	Lys	Ser	Gly	Ala	Gly	Glu	Arg	Gly	Ser	Arg	Gly	Ile	Ile	Ala	Ala		
						210			215			220					
Leu	Gly	Pro	Asp	Gly	Lys	Pro	Ser	Arg	Ile	Val	Val	Ile	Tyr	Thr	Thr		
						225					235				240		
Gly	Ser	Gln	Ala	Thr	Met	Asp	Glu	Arg	Asn	Arg	Gln	Ile	Ala	Glu	Ile		
						245			250					255			

FOUO 44-01250

Gly Ala Ser Leu Ile Lys Trp
260

<210> 10
<211> 262
<212> PRT
<213> Bacillus licheniformis

<400> 10
Met Lys Asp Asp Phe Ala Lys Leu Glu Glu Gln Phe Asp Ala Lys Leu
1 5 10 15
Gly Ile Phe Ala Leu Asp Thr Gly Thr Asn Arg Thr Val Ala Tyr Arg
20 25 30
Pro Asp Glu Arg Phe Ala Phe Ala Ser Thr Ile Lys Ala Leu Thr Val
35 40 45
Gly Val Leu Leu Gln Gln Lys Ser Ile Glu Asp Leu Asn Gln Arg Ile
50 55 60
Thr Tyr Thr Arg Asp Asp Leu Val Asn Tyr Asn Pro Ile Thr Glu Lys
65 70 75 80
His Val Asp Thr Gly Met Thr Leu Lys Glu Leu Ala Asp Ala Ser Leu
85 90 95
Arg Tyr Ser Asp Asn Ala Ala Gln Asn Leu Ile Leu Lys Gln Ile Gly
100 105 110
Gly Pro Glu Ser Leu Lys Lys Glu Leu Arg Lys Ile Gly Asp Glu Val
115 120 125
Thr Asn Pro Glu Arg Phe Glu Pro Glu Leu Asn Glu Val Asn Pro Gly
130 135 140
Glu Thr Gln Asp Thr Ser Thr Ala Arg Ala Leu Val Thr Ser Leu Arg
145 150 155 160
Ala Phe Ala Leu Glu Asp Lys Leu Pro Ser Glu Lys Arg Glu Leu Leu
165 170 175
Ile Asp Trp Met Lys Arg Asn Thr Thr Gly Asp Ala Leu Ile Arg Ala
180 185 190
Gly Ala Ala Ser Tyr Gly Thr Arg Asn Asp Ile Ala Ile Ile Trp Pro
195 200 205
Pro Lys Gly Asp Pro Val Gly Val Pro Asp Gly Trp Glu Val Ala Asp
210 215 220
Lys Thr Val Leu Ala Val Leu Ser Ser Arg Asp Lys Lys Asp Ala Lys
225 230 235 240
Tyr Asp Asp Lys Leu Ile Ala Glu Ala Thr Lys Val Val Met Lys Ala
245 250 255
Leu Asn Met Asn Gly Lys
260

<210> 11
<211> 30
<212> DNA
<213> Drosophila melanogaster

<220>
<221> misc_feature
<222> (0)...(0)
<223> n = A, C, T, or G

<400> 11
ntntctctct tttctctctc tctcncaggt

30

<210> 12
<211> 93
<212> DNA

097244-012601

<213> Artificial Sequence

<220>

<223> Truncated En-2 splice acceptor

<400> 12

caacctcaag ctagcttggg tgcgttggtt gtggataagt agctagactc cagcaaccag
taacctctgc ctttctcct ccatgacaac cag

60
93

<210> 13

<211> 10

<212> DNA

<213> Artificial Sequence

<220>

<223> Splice donor sequence

<221> misc_feature

<222> (0)...(0)

<223> n = A, T, G, or C

<221> misc_feature

<222> (0)...(0)

<223> r = A or G

<400> 13

nagggttagt

10

<210> 14

<211> 10

<212> DNA

<213> Artificial Sequence

<220>

<223> Splice donor sequence

<400> 14

gaggtaagta

10

<210> 15

<211> 15

<212> DNA

<213> Artificial Sequence

<220>

<223> Splice donor sequence

<400> 15

caggtgagtt cgcac

15

0977244-012604